DOCUMENT RESUME

ED 340 763 TM 018 019

AUTHOR	Ratcliff, James L.
TITLE	What Type of National Assessment Fits American Higher Education?
SPONS AGENCY	National Center for Education Statistics (ED), Washington, DC.
PUB DATE	Nov 91
NOTE	46p.; Commissioned paper prepared for a workshop on Assessing Higher Order Thinking & Communication Skills in College Graduates (Washington, DC, November 17-19, 1991), in support of National Education Goal V, Objective 5. For other workshop papers, see TM 018 009-024.
PUB TYPE	Viewpoints (Opinion/Position Papers, Essays, etc.) (120) -- Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Cognitive Measurement; College Faculty; *College Graduates; Communication Skills; Critical Thinking; *Educational Assessment; Educational Environment; Educational Improvement; Educational Policy; *Evaluation Methods; Higher Education; *National Programs; Problem Solving; *State Programs; Student Evaluation; Test Construction; Thinking Skills
IDENTIFIERS	National Education Goals 1990

ABSTRACT

For a national program assessing college graduates to succeed, it must be institutionalized at the majority of colleges and universities in the nation. The information generated must play a clear and key role in formulating state and institutional higher education policy, college matriculation standards, and degree-granting decisions. Such an assessment must be particularly credible, reliable, and useful to college faculty, who play a fundamental role in communicating the importance of the assessment program to students. Such confidence can only be developed through the active participation of institutions and faculty. A national effort should build on the credibility and accomplishments of existing state programs, rather than being duplicative and ancillary. A two-pronged approach to development of a national assessment program seems warranted. For the short term, protocols to monitor selected course syllabi and examinations can be developed. A long-range goal would be adoption of a national assessment program. Using the methodology and model from the Differential Coursework Patterns project, the National Center for Education Statistics could begin to determine which measures of critical thinking, communication, and problem solving best differentiate between appropriate and inappropriate learning environments for students of different ability levels. A 21-item list of references is included. Reviews by N. Beck, J. Herman, and T. Marchese and B. Wright of this position paper are provided. (SLD)

ED340763

WHAT TYPE OF NATIONAL ASSESSMENT FITS AMERICAN HIGHER EDUCATION?

James L. Ratcliff

- Abstract -

For a national program assessing college graduates to succeed, it must be institutionalized at the majority of colleges and universities in the nation. The information generated from such an assessment must play a clear and key role in formulating state and institutional higher education policy, in college matriculation standards and degree-granting decisions. In short, it must affect the teaching of faculty and the learning behavior of students.

For colleges and universities to give such meaning to an assessment program, it must be credible, reliable and useful. It must be particularly credible, reliable and useful to faculty, who play a fundamental role in portraying the importance of the assessment program to students. Such confidence can only be developed through the active participation of colleges and universities and their faculty in the development of such an assessment program. Several states and many institutions have substantial investments in currently operating and successful assessment programs. The national effort should build on the credibility and accomplishments of these programs rather than to be duplicative and ancillary to them.

A two-prong approach to the development of a national assessment program seems warranted. For the short term, protocols could be developed to monitor selected course syllabi and examinations to determine the extent to which they encourage the development of communications, critical thinking and problem-solving abilities. A longer range goal would be the adoption by states and institutions of a national assessment program. Using the methodology and model from the Differential Coursework Patterns Project, NCES could begin now to determine which measures of critical thinking, communications and problem-solving best differentiate between appropriate and inappropriate learning environments for students of different ability levels.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM018019

BEST COPY AVAILABLE

WHAT TYPE OF NATIONAL ASSESSMENT FITS AMERICAN HIGHER EDUCATION?

James L. Ratcliff

Director,
Center for the Study of Higher Education
The Pennsylvania State University
and
Co-Director,
National Center for Postsecondary
Teaching, Learning and Assessment

Over the past decade there has been a rising concern over the quality of American education. While this concern has been mainly with the elementary and secondary grades, certain quality issues pertain directly to higher education. The higher education reform proposals of the 1980's made implicit assumptions about what constituted effective undergraduate education. The three most frequently cited of these, Involvement in Learning (NIE, 1984), To Reclaim a Legacy (Bennett, 1984), and the American Association of Colleges' Integrity in the College Curriculum (1985), cited a decline in the quality of liberal or general education and called for reforms to strengthen undergraduate programs. The NIE report called for more active student involvement in the learning process. The NEH Bennett report called for the restoration of the humanities to a central position in the undergraduate curriculum. The AAC report called for a redefinition of the meaning and purpose of baccalaureate degrees. Despite a spate of curricular reform in the late 1980s on many U.S. campuses, undergraduate education remains a diffuse, discursive and expanding array of coursework and programs. Now the National Educational Goals Panel has called for "the proportion of college graduates

who demonstrate an advanced ability to think critically, communicate effectively, and solve problems" to increase substantially.

Like the national reports that recommended higher education standards be raised, the curricular prescriptions for students be changed, and the content and structure of degree requirements be fortified, the objective of the National Educational Goals Panel (NEGL) relative to college graduates assumes that there is a common body of knowledge, skills and abilities to be imparted through an undergraduate education and that there are readily available means to assess student progress in acquiring that common knowledge. Yet, our current ability to make such assessments is limited by the relative lack of common curriculum and relatively limited range of means and measures of that curriculum. In this paper I will highlight what are the obstacles to assessing students' critical thinking, clear communications and problem solving abilities and propose ways that we can move rapidly to better understand what students gain from college in these three important areas of learning. To begin, I first examine the initial impetus to assessing collegiate level learning.

Factors Affecting the Success of a National Collegiate Assessment Program

Several factors have lead American higher education to adopt forms of quality control based on assessment in the past. These factors including a) wide variation in the quality of secondary education, b) rapid increases in college admissions among people previously unexposed to higher education, and c) rapid change and expansion in the

curriculum. These three factors combined in the past to lead colleges and universities to more effectively monitor the quality of their students. These factors again affect today's efforts to monitor and measure students' abilities to think critically, communicate clearly and solve problems. Each force has a fundamental role in shaping the development of an effective national program of collegiate assessment.

Higher education has always had unique responsibility as a standards bearer within the American educational system. This is because colleges and universities were established usually before a system of secondary education developed. Harvard was founded long before there was any widespread college preparatory programs. Land-Grant colleges and universities were established long before universal secondary education was extended to the rural areas of this country. During their first years of operation, often more than half of the students at Land-Grant colleges were in precollegiate studies. Women's colleges, historically Black colleges and universities, colleges and institutes for American Indians all were established before there were secondary educational programs to prepare these groups for collegiate level studies. This curious historical phenomena consistently has placed higher education in the position of judging the qualifications of the students it admits, thereby articulating academic standards for college preparatory and secondary education in the process. It has also given higher education the responsibility for providing precollegiate instruction to remediate those students without adequate levels of academic preparation to succeed in college.

While the NEGP's primary concern is with the abilities of college graduates, this

concern is directly related to the quality of educational preparation of students admitted to college. College admissions standards establish the starting point for a college education, the beginning criteria against which the value added by a baccalaureate degree may proceed and should be judged. Also, the adoption and nearly universal use of standardized testing as a primary basis for college admissions also contains some lessons for those who seek to monitor the development of college graduates.

The poor quality of secondary education at the beginning of the twentieth century was a prime motive for colleges and universities to establish the College Entrance Examination Board (CEEB) and to adopt its testing program as an admissions criteria. It should be noted here that it was the failure of various states to generate meaningful educational policies and adequate support for secondary education which motivated higher education to adopt the CEEB testing program. For example, at the beginning of the century California had less than one public high school per county and had a constitutional provision which forbad the use of state funds in providing secondary education. The CEEB initiative to provide uniform college admissions requirements and examinations came from the Association of Colleges and Secondary Schools of the Middles States and from the leadership of president Nicholas Murray Butler of Columbia University. The CEEB program was *not* imposed on colleges and universities by state or federal government. Rather, the need was widely seen by the college leaders of the day, thereby enforcing its widespread acceptance and success in higher education (Brubacher and Rudy, 1976). The CEEB *did* dictate standards for secondary education and for student performance in key areas of knowledge, skill and ability.

While they is widespread concern over the academic preparation of students entering higher education, there is a lack of consensus as to the exact nature and extent of the problem. Similarly, there is disagreement as the strengths and deficits of contemporary undergraduate education as manifest in the abilities of today's college graduates. There are multiple visions of what constitutes intelligence and learned abilities (Ratcliff, 1990; Sternberg, 1989). Enlightened approaches to assessment include multiple definitions of learned abilities and multiple measures of that learning.

Such approaches avoid the essentialistic quest for the one best set of measures that will encompass all of general learning and cognitive development at the collegiate level. So long as we believe that studying different subjects produces different types of learning, and so long as higher education forwards a curriculum that attempts to embody the expanding diversity and complexity of advanced human thought, multiple definitions of the ingredients to intelligent behavior and multiple means to assess them will be required. Not all diversity is good, however. The current divergence in the quality of preparation of high school students severely inhibits colleges and universities' ability to foster higher levels of skill in critical thinking, communications and problem solving. A fundamental aspect of an effective national assessment of college graduates involves the assessment of students as the enter college. Since a growing number of these students do not come directly from high school and are working adults, such a program cannot rely alone on assessment programs based in the secondary schools.

A second great force between the original adoption of the CEEB testing program was rapid rise in immigrants in the first two decades of this century. The educational

needs and the educational backgrounds of students diversified rapidly. The influence of the suffrage movement and the growing educational expectations of women encouraged coeducation and collegiate education for women and further contributed to the diversification of students applying for admission to college. At the outset of the decade, less than four percent of the American population (238,000) went to college. By the end of the 1920s, 12 percent of the high school graduates were attending college. Since 1980 there has been 6 million new immigrants to the United States. Once again we are expanding and extending higher education to new segments of our population. Some come with little or no formal education or language skills, others come with extensive education but few language skills, while a third group consist of those with English language skills but little formal education. Most of these new arrivals are outside the normal elementary to secondary to higher education scheme. For example, a survey of English as a Second Language (ESL) students at San Jose City College revealed *in the fall of 1981*, 86 percent were Vietnamese, *with the remainder coming from 10 other countries*, 79 percent had been living in the United States for 2 years or less, 65 percent spoke Vietnamese and 21 percent spoke Chinese, 90 percent were refugees, and 21 percent had more than 12 years of education in their home countries. *By the fall of 1987*, only 35 percent of the ESL students were Vietnamese, *with the remaining students coming from 76 different countries*, 63 percent had lived in the United States 3 years or longer (Gosak, 1988).

As the SJCC example suggests, the diversity of cultures and educational backgrounds of college students is expanding rapidly. The federal government has

encouraged this expansion through incentives to colleges who serve such groups as displaced homemakers, students with disabilities, those needing adult basic education, and the unemployed seeking job retraining. Programs targeted for these students have broadened both the curriculum and the demographic profile of students served by it. In the 1970's and at the outset of and 1980's, many higher educators predicted enrollment declines based on the decline in secondary school enrollments. These declines failed to materialize. Instead, the demand for higher education rose as the value of a high school education declined. Much of this new enrollment did not come from the traditional 18-to-24 year old college-going cohort. It was the newly-arrived immigrant, the working adult, the mother reentering the workforce, and the military veteran that helped swell and diversify enrollments.

The diversification and expansion of the entering freshmen meant that the quality, relevance and recency of their secondary education would be far less uniform and far less ascertainable. Colleges and universities enrolling these students are less likely to have current records of academic ability or achievement. Similar to the educational challenges of the turn of the century, we once again face a pressing need to set clear standards for the articulation of secondary, precollegiate and higher education.

A third force in the diminution and divergent in quality of the undergraduate experience has been the effect of the explosion of knowledge itself, reflected in unbridled expansion of the college curriculum. Since the adoption of free electives at Harvard in the 1890s, the undergraduate curriculum at most colleges and universities has continued to grow to incorporate new courses and new programs. Area studies, women studies,

computer science and nuclear physics are but a few of the new subjects of this century. As the curriculum expanded and as coursework was added, coherence and purpose to the baccalaureate were increasingly difficult to discern (Veysey, 1973). The curriculum of the medieval university had sequencing and prerequisites. The subjects of the quadrivium (arithmetic, geometry, astronomy, and music) were to precede study in the trivium (logic, grammar, and rhetoric). Such notions of development and sequence have been largely replaced by the more oblique notion of breath and depth of study in undergraduate study. The evidence has grown that students at the same institution do not share a common curricular experience in their pursuit of the bachelors degree (Boyer & Ahlgren, 1987; Ratcliff, 1990).

We do need a means for determining students' ability to think critically, communicate clearly, and solve problems. To accomplish this goal, it has been suggested that the National Center for Educational Statistics develop a collegiate-level test similar to the National Assessment of Educational Progress (NAEP). However, the call for a NAEP-like national examination does not build on any widespread institutional and faculty sense that such testing would improve undergraduate education. The NEGP call presumes that commonly accepted means and methods are available to assess student growth in critical thinking, communications and problem solving abilities. The call presumes not only a consensus of criteria on what constitutes learned abilities in these three areas but also assumes that, given the creation of an appropriate yardstick, higher education will have the means and the resources to improve student performance. In short, the current call for a NAEP-like examination fails to address fully the three

aforementioned factors which have largely contributed to the need and call for such an assessment.

Higher education is different from elementary and secondary education in ways that are likely to doom the immediate implementation of a national collegiate testing program to failure. There are salient features of collegiate education which mitigate against such a testing program as the most effective means to accomplish the NEGP goal. There are alternative strategies which are more in keeping with the policy, governance and funding structures of American higher education, with the academic culture that has spawned the most sought-after higher education system in the world, and are more likely to sustain and enhance our commitment to social democracy and public access to higher education.

Students and Institutions Must Be Motivated to Participate in the Assessment Program

Because higher education is voluntary, the level and quality of student participation in any national testing program are significant issues. Similarly, the personal importance and significance of the testing to the individual student also affects that student's motivation to take the test and to make a serious effort to complete it to the best of that individual's ability. In short, there must be sufficient student motivation to participate in any testing program in order to achieve valid and reliable results. Students must cooperate for the test results to be meaningful (Borg and Gall (1983). Two important considerations in encouraging student participation

include the significance of the tests to the students personally and the significance of the tests to the institution.

The SAT and ACT tests have significance to students because they have bearing on the student's ability to enter college. Similarly, the GMAT, MCAT, LSAT, GRE, and Miller's Analogies tests have importance for the minority of college graduates who plan to go on to graduate or professional studies. This is because the tests have significance to both the institution and the individual. Such meaning results from both the institution and the individual recognizing the assessment program as an integral, legitimate and important part of the curricular and matriculation procedures.

We stand in real danger of reversing our accomplishments in providing access to higher education by a casual and cavalier implementation of a national assessment program. Consider the experience in Florida. Given the current standards for Florida's College Level Academic Skills Test (CLAST), only 48% of the first-time CLAST examinees are expected to pass the test, and the impact on minority students is even more severe. The standards are based on the assumption that improvements in Florida's high school education during the 1980's would raise graduates' academic competence and skills base. A similar assumption undergirds the quest for a national college-level NAEP-like exam. Yet, over the past 10 years, the percentage of Florida students enrolling at Miami-Dade Community College with skills deficiencies has remained between 61% and 66%. After two years of community college education, few of these students have been able to improve their competencies enough to pass the CLAST exam. While the intent of the CLAST was to improve the ability levels of

students entering Florida colleges, the consequence has been a significant decrease in the number of individuals completing college degrees in Florida. Those who do graduate are able to pass the College Level Academic Skills Test (CLAST) at a considerably higher level (McCabe, 1990). Similar findings are reported relative to the Texas Academic Skills Program (TASP) which prevents students from enrolling in more than 10 credits of college-level coursework without passing the basic skills examination (Smith & Morris, 1989). The Florida and Texas assessment programs give the institution and the individual student a clear stake in the outcome of the test. Yet, these state policies may have traded away educational access and opportunity for educational achievement on a standardized test. High school students have not improved their knowledge, skills and abilities in these states. Instead, fewer students with higher entering abilities go on to college. Mandating testing will not alone improve education, and alone it may well reverse educational policy in the process.

Tests must have diagnostic and/or predictive value to an institution. The tests must have bearing on college entrance, continuance or graduation for students to take them seriously. Without such belief in the testing program by college administrators and faculty and without some value to the student in participating in the testing program, the validity and reliability of the results will be continuously suspect. Tests are not inherent deterrents to enrollment. When they are perceived as fair and important, they are accepted and students complete them to the best of their ability. A study of competency testing in North Carolina and Louisiana suggested the majority of students, most of whom were black, believed the tests were fair and necessary, and

included content that they had studied. However, these students did not believe that the tests stimulated them to study harder nor did they think it caused their instructors to have higher performance standards nor did it alone cause them to change their course selections (Reed, 1987). In short, such an assessment program alone cannot bring about improvement in learning.

There can be a significant diagnostic value of testing to the individual student as well. Statewide assessment programs have been useful and beneficial to student learning in California (Swordes, 1990), Colorado (Richards 1986), and Tennessee (Hobbs, 1989) when students used the test information to choose coursework which was challenging and appropriate to their ability level. The major differences in student abilities among students on the same campus rather than between students from different campuses. The real challenge is to use curricular and assessment information to help student select coursework that is appropriate to their interests, abilities and skills. Assessment in service of finding proper matches between student abilities and educational environments are far more likely to produce gains in student learning than are those that seek to discourage and impede the progress of the less able student.

Involvement in Learning (1984) recognized the importance of using faculty in the assessment of student learning. A first step in that process is to establish the credibility of the assessment program in the eyes of college faculty. A second step is to find ways to involve them directly or indirectly in the assessment process. In doing so, the assessment program also insures that faculty understand fully what is

being assessed. Faculty are in the best position to draw students' attention to the outcomes their college or university values the most. As the Involvement of Learning report indicates, assessment has even greater potential as a tool for clarifying expectations and for increasing student involvement when it is used to measure improvements in performance.

An example of how faculty and institutions can be encouraged to incorporate an assessment program in the teaching, learning process is the California Basic Skills Instruction policies of the California Community College system. The California Precollegiate Basic Skills Instruction policies require colleges to establish ability level prerequisites for degree-applicable, entry-level courses. These policies require institutions to develop standards for the rate at which a student progresses toward a degree and limits students to earning a maximum of 30-semester-units in precollegiate basic skills courses. Each institution must define the scope of their student assessment program and relate them to student course selection. A statewide task force examined the costs of the implementation of these policies to the individual institutions and to the state (Farland & Cepeda, 1990). The Basic Skills Instruction policies are state-level reforms accompanied by recommendations for the fiscal support and curricular change necessary to affect the purposes of the program. In short, the role of assessment on these policies was a means to an end, not an end in itself.

Unlike secondary education, college attendance is voluntary. As a country we take pride in noting that nearly half of the high school graduates in the United States go

on to college. As the impact of the Florida CLAST and Texas TASP suggests, add-on assessments of student learning can discourage college attendance and college completion among non-traditional students.

There are few demonstrably successful testing programs that include representative proportions of part-time students, working adult students, minority students in commuter institutions. What if the nation gave a test and only a few students showed up to take it? At the National Center for Postsecondary Teaching, Learning and Assessment, we are examining ways to encourage working adults and other non-traditional students to participate in a comprehensive, longitudinal assessment program that includes measures of critical thinking and communications abilities. We are cognizant that a national testing program may dissuade or exclude the very individuals for whom the national objective of increasing critical thinking, communication and problem-solving skills may be most needed. As an admissions' test, the Scholastic Aptitude Test has been repeatedly attacked for alleged bias against women, African American students, American Indian students. By creating a new test, we are not likely to produce a measure that is less assailable to these charges.

Some argue that admissions testing is not a necessary part of the strategy to monitor and demonstrate improvement in critical thinking, communications and problem-solving abilities of college graduates. Yet, what good is an assessment program that fails to inform us in ways useful to the improvement of collegiate education? If today we had a national test of critical testing administered at the conclusion of the baccalaureate degree, we would be able to tell if students in any one year did better or

worse than students in preceding years. Let us suppose for a moment that we had such information for several years and the statistics suggested a steady decline in the students' ability to problem-solve. Without parallel admissions testing, we would be unable to determine if the decline was attributable to secondary or postsecondary education. The journalists' pension to administer guilt by association would no doubt lead several reporters to assume that colleges were directly responsible for such declines.

We know that in years of economic recession and depression more people, particularly unemployed adults and women returning to the workforce enroll in college. As the proportion of the population that attends colleges increases, the averages of student scores on ability and achievement tests decline. Thus, as colleges and universities serve a broader array of student abilities, the proportion of academically-talented students is likely to grow smaller. A national assessment program needs to be able to distinguish between a decline in the quality of educational programs and an increase in student participation in higher education.

If parallel testing of critical thinking, communications and problem-solving abilities were administered in the high schools, there would be baseline to evaluate the extent to which colleges contribute to students' development of these abilities. However, this baseline information would only be available for those students who we already know the most about, who are most likely to perform well on academic achievement tests, and who are best prepared to succeed in college. We miss gathering information on the very students that we strive to encourage, include and provide access to higher education. We will not have a clear picture of the extent to which students improve in these key skills

if our information includes only part of the college-going population.

A frequent rationale for a collegiate-level testing program is that it will help insure our global competitiveness by having better educated workers. It should be noted that in other industrialized countries with whom we compete, including Germany, France, Great Britain and Japan, national testing programs work to exclude all but a small proportion of students from going to college. College-going rates in these countries range from 17 to 25 percent. Great Britain and France currently are seeking to expand access to higher education with the goal of doubling the college-going rate. They acknowledge that matriculation examinations often currently work to limit participation in higher education. It is no small irony that our competitors in the industrialized world are seeking to create more open and accessible higher education systems at a time when we seek to contract, exclude and be more selective. Surely global competitiveness cannot be further by both greater selective and greater access. Are we on the right path to global competitiveness? We desperately need to examine what we mean by that term before accepting it blindly as a rationale for educational reform.

The NAEP testing program uses matrix sampling to achieve a profile of student achievement with a minimum of obtrusiveness and expense. Elementary and secondary curriculum are relatively uniform. Some states have uniformly adopted or approved textbooks. Curriculum guides exist for a relatively fixed range of subjects. School principals regularly monitor teacher lesson plans and course materials for consistency of purpose and method. There is an overall unity to the elementary and secondary

educational programs that makes it amenable to testing.

No such uniformity exists in higher education. A review of the last 20 years of research on the effects of college on students indicates fairly clearly that the differences in student learning within colleges and universities is greater than the differences between colleges and universities (Pascarella and Terenzini, 1991). Students enrolling at a modern university may have 3,000 to 5,000 courses from which to choose the 35 to 45 courses they will take to complete their bachelors programs. Students at a small college may have several hundred such courses from which to choose. The curricular uniformity characteristic of secondary education is not present at the postsecondary level.

States and Students should be the Focus of Assessments

Most colleges and universities have distributional educational degree requirements. This means that students typically have a wide range of course choices to make in order to complete degree requirements. Students in the same graduating class may have very few courses that they have all taken. With such diversity in the formal educational program, is it any wonder that the differences in learning are greater within institutions than between them? In reality it is the students who select their educational program, and that educational program is what is to be evaluated in terms of its development of critical thinking, communications and problem-solving skills.

This is a very important point for any national assessment program. For an assessment to be effective, it must not simply tell us how well we are doing, but it must also do so in such a way that we know how to improve on the situation. Let me return

to the instance of the university with 5,000 courses in which undergraduates may enroll. Let us suppose that our assessment shows an overall decline in problem-solving ability. To improve the educational program, it is a foolhardy waste of resources to insist that all 5,000 courses increase their focus on problem-solving. First, not all students in the group tested took all the courses. Secondly, some students presumably improved while many did not. An effective assessment program will tell us which educational sequences of coursework lead to the improvement of these learned abilities and which did not (Ratcliff, 1990). Without such differentiation we a merely slinging mud at the institution.

A second consequence of the finding that learning differences are greater within institution than between them is that the unit of analysis should not be the institution. Yet, most of the assessment and reporting efforts have been at the institutional level. Each institution attracts a different student population. If we had only a senior-level exam in critical thinking, communications and problem-solving, clearly the selective state university would appear to be better that the local community college. The test results would only report on the differences in admissions criteria and not on the educational programs themselves. The testing who provide little guidance to the faculties of each institutions as to which programs were helpful, which needed strengthening, and why. The assessment program would fail because it would not provide reliable, valid information.

Colleges and universities were often created or chartered by state legislatures as the result of lobbying by one or more interest groups. Funding of these institutions is

derived of partisan politics. Rather than crafting meaningful educational policy, legislatures have largely been decided on a year to year basis which colleges or college systems should get how much of the state budget. Rarely has the question been asked in the budgetary process as to what the sum of higher education institutions serves to provide the state. This focus of single institutions and systems of institutions has led assessment efforts to commit two logical errors corrupting the quality of the information. First, the reductionist error is regularly committed in interpreting assessment data. Grant and Sleeter (1986) noted, for example, that attention to one student group (blacks, women, etc.) oversimplifies the analysis of student behavior, reducing that behavior to the group stereotypes. We need to understand better what is an assessment of critical thinking, communications, or problem-solving ability that clearly differentiates among those women and minorities who succeed and profit by their collegiate experience and those who do not. A second error is the generated from arguments asserting the uniqueness of data coming from a single institution (or institutional type) of higher education. Community colleges, for example, might argue their missions are different from other higher education institutions, suggesting that the results of an assessment cannot be generalized to them in the same manner as research universities. However, such an argument fails to acknowledge that community college faculty come from the same socialization process that characterizes four-year college faculty. The purchase textbooks from the same book dealers, they participate in the same discipline-based associations, and their efforts to articulate coursework from junior to senior level institutions invariably link them to other higher education institutions. Institutions of

higher education exist formally and informally as systems nested within and linked to larger educational systems and networks and containing smaller subsystems (departments, divisions and programs) that are in turn linked to them. Assessment of student critical thinking, communications and problem-solving abilities is an investigation conducted within institutions of higher education but not necessary of them.

In the United States, it is the states that are charged to establish the schools, not the federal government. Most of the financial support for all levels of education comes from state and local government, not national. It is the states who charter private and independent colleges and who maintain a responsibility for accrediting institutions of higher education. If an educational report card is to be issued, then it should provide between state comparisons rather than interinstitutional comparisons. States set the standards for the educational programs that ready students to enter college. States decide on the level of support to provide public and independent forms of higher education. States have varying policing regarding the provision of higher education for minorities, women and adults. This year-to-year change is a focal point of the national longitudinal study of first year college students planned to commence next fall at the National Center for Postsecondary Teaching, Learning and Assessment.

We Need to Discover which Learning Environments Benefit which types of Students

The question is clearly not "Which colleges provide a better education?" Such a question holds little meaning to anyone other than those that think that there is but one

standard for quality throughout higher education. Instead, the guiding question for assessment should be, "Which groups of students benefit most from which collegiate environments?" Only by answering the second question can we hope to show substantial increases in students' ability to reason critically, communicate clearly or to solve problems. A national assessment should identify between state differences and between student differences. The between student differences could be defined according to academic ability or through focused study of groups with low participation or success rates in college.

Some suggest that the value added to a student's abilities might be determined by administering a test on admission to college and upon graduation. The difference in the pre-test and post-test scores might show learning acquired during college. While such a design is clearly superior to simply evaluating student achievement upon graduation alone, it does not account for the substantial difference between the normative groups for the freshman and senior exams. Regrettably only half of the individuals who enter American colleges and universities ever complete the baccalaureate degree. Thus, the normative standards on the senior examination would be by necessity higher than those of the freshman exam (Adelman, 1988). The greatest proportion of attrition occurs in the freshman year of college. Ideally, an assessment of student abilities in critical thinking, communication and problem-solving would occur during all four years of college so that the year-to-year contribution of the educational program could be described. Only through such detailed information can the planners of the curriculum and extracurriculum plan meaningful activities.

Not only is there significant variation in learning within a single institution, but also there is significant year-to-year variation in student learning at that institution. Due to the expansiveness of the curriculum, the course availability in any given term of enrollment, and the variation in student selection of coursework, students in one graduating class may experience up to a 25 percent difference in coursework from those in the preceding or following years (Ratcliff, 1990). This variation in learning experiences varies the extent and type of knowledge, skills and abilities students acquire and their performance on test batteries. This variation does not necessarily portray year-to-year dips and rises in graduating student abilities. Rather, it may be simply a function of the courses and subjects they choose to study. When a university offers an undergraduate the opportunity to pick 35 to 45 courses from a curriculum of 3,000 to 5,000 courses to complete the baccalaureate, it is little surprise that different college graduates evidence different levels and types of knowledge, skills and abilities. The challenge to an effective assessment program is to cast its definitions of what constitutes clear communications, critical thinking and problem-solving sufficiently broadly to capture the full range of learning associated with these terms.

Assessments of students' critical thinking abilities illustrate this point. One factor differentiating tests of critical thinking is that of problem structure. Problem structure is the extent to which a problem can be described fully and can be answered rightly or wrongly. Complex social, political or economic problems do not have right or wrong answers. Often their very nature is debated. These are ill-structured problem sets. In contrast, problems that can be solved by deductive logic (in the spirit of Sherlock Holmes

or Miss Marple) possess a high degree of certainty and correctness. They are well-structured problems.

Two popular measures of critical thinking are the Cornell Critical Thinking Test (CCTT) and the Watson-Glaser Critical Thinking Appraisal (WGCTA). Each measures a student's ability to solve structured problems. Each has strong correlations with the ACT, SAT and GRE examinations. For example, King, Wood & Mines (1989) found that the WGCTA correlated with the ACT at $r^2=.59$ and the CCTT correlated with the ACT at $r^2=.62$. If sixty percent of the variance in scores on a college-level critical thinking test is explained by a traditional, standardized measure of student achievement, what is the critical thinking test really measuring? The difference between what standard tests of general learning measure and what certain tests of critical thinking assess may not be fundamentally different. In adopting multiple measures to assess general learned abilities among colleges students, researchers have an obligation to decide if each measure evaluates separate and distinct attributes of learning.

Students in the Differential Coursework Patterns (DCP) Project at the National Center showed significant improvement on the Analytic Reasoning (ARE) and Logical Reasoning (LR) item-types of the Graduate Record Examination (Ratcliff, 1990). To what extent do these GRE item-types represent critical thinking or higher order reasoning skills? We need further research to identify the types of knowledge and cognitive abilities required to answer these type questions. To what extent do they, for example, suffice a college's need (or a nation's need) to measure the development of critical thinking abilities among its students?

The DCP Project research suggests that the item-types on currently used standardized tests (such as the ACT, SAT and GRE) can be used effectively to describe specific general learned abilities. This conclusion held true for the broad and representative spectrum of graduating seniors examined in the Project. These seniors came from both selective and open-admissions institutions of higher education. High ability students who showed significant improvement in one or more general learned abilities enrolled in different coursework patterns from those students of low entering ability who showed comparable score gains. The research showed that there were different effective coursework patterns for students of different entering abilities. There was little evidence to suggest that the GRE test failed to differentiate between the pre- and postcollege achievement of high ability students.

Our Current Means of Evaluating Student Learning is not Satisfactory for a National Collegiate Assessment

Higher education curriculum has always been viewed as a developmental sequence of learning events, one course building upon the learning imparted by the preceding ones. Essentialist and constructionist theories of curriculum stress the importance of combinations of subjects (core curricula, great books, etc.) as influential on general learning. Behaviorists emphasize the process of learning and focus on skill development. Still, they too see the logical sequence of tasks, from the simple to the complex, as a necessary element in the learning process. The medieval university curriculum was organized according to combinations and sequences of courses as well as individual

subjects. The seven liberal arts were organized in an ascending hierarchy of prerequisites.

The point is that most theories of curriculum call for a developmental sequence of learning in order to produce a result. The results we are concerned with in this instance are critical thinking, communications, and problem-solving abilities. We are not interested in determining the sum of all such learning experiences, nor the average performance of students in learning such abilities, but rather the effectiveness of the progression of learning in producing the desired results.

It is the search for a more efficacious curriculum that leads us to assessment. We have a means of judging student learning which we widely accept and use: course grades. Course grades, however, vary from subject to subject, from professor to professor, from class to class of students. The Grade Point Average (GPA) tells us how well a particular student performed relative to that students' class peers. It gives us no broader normative basis for comparison. More importantly, it provides us with no direct information on the progressive or cumulative development of specific learned abilities such as problem-solving, critical thinking, or communications.

What Can Be Done

A program to assess the critical thinking, communications and problem-solving abilities of college graduates should logically start in the colleges themselves. This beginning should proceed under state educational policy rather than direct but ancillary intervention of the federal government. A two-prong approach seems warranted.

For the short term, protocols could be developed to monitor selected course

syllabi and examinations to determine the extent to which they encourage the development of communications, critical thinking and problem-solving abilities. We at the National Center for Postsecondary Teaching, Learning and Assessment are developing such protocols for analysis. The Center for Assessment Research and Assessment at the University of Tennessee is developing standards of good practice that could also be applied to classroom assessment practices as well. Institutions could be encouraged to develop these indirect indicators of student learning in the three key areas of assessment. These indicators would not only provide initial estimates of how students are progressing as well as the extent to which current college curriculum is directed to the enhancement of critical thinking, clear communications and problem-solving abilities.

A longer range goal would be the adoption by states and institutions of a national assessment program. Using the methodology and model from the Differential Coursework Patterns Project at the National Center, we could begin now to determine which measures of critical thinking, communications and problem-solving best differentiate between appropriate and inappropriate learning environments for students of different ability levels. Rather than become embroiled in debate over what constitutes critical thinking or clear communications, we could begin an investigation of what existing measures overlap each other, which best describe student improvement, and which are most closely aligned with the curriculum of particular institutions or student groups. This second prong of investigation would move us closer to understanding how we may use assessment information to improve students' abilities in these key areas articulated in the NEGP objectives. We suggest three applications and extensions of the

DCP Project research and its analytic model:

The purpose of the "Differential Coursework Patterns (DCP) Project" was to determine the effect of different patterns of college coursework on the general learned abilities of students. To accomplish this end, a model for linking what coursework students took in college with what they learned in college was developed. The result was the Cluster Analytic Model. The Model groups courses appearing on student transcripts according to the distribution of assessment scores of those students. The Model uses precollege indicators of student learning to control for incoming student ability. It uses transcripts, rather than formal course or degree requirements as the representation of the college curriculum. The Model can use any number of assessment measures, including both quantitative and qualitative data.

What does the DCP Cluster Analytic Model tell us about assessment? The Cluster Analytic Model uses multiple measures of assessment. It provides colleges with information regarding the extent of variation in student assessment results that is explained by any one of the measures used. This information can be helpful in a number of ways. Faculty and administrators need not decide on an ideal set of assessment measures. The extent to which such measures may overlap in describing student learning can be identified. The mix of assessment measures appropriate to the goals of the college and the characteristics of the student population can be continuously monitored. When students show small amounts of growth on an indicator of student learning, either the college can develop strategies for improving student learning in the area identified, or discard the measure as inappropriate to the college and its students. The Cluster

Analytic Model provides useful information to the college about the mix of assessment measures that reflects what the students learn and what the college intends to teach them.

What does the DCP Cluster Analytic Model tell us about the curriculum? The Cluster Analytic Model is a tool ideally suited to institutions of higher education with a distributional general education requirement and a wide array of programs, electives and majors. For example, if one of the assessment measures a college selects is a test of analytic reasoning, then the Cluster Analytic Model can identify those groups of courses that students took who showed significant improvement in that area of general learning. Furthermore, the student population can be subdivided into high ability and low ability students, by gender, race or ethnicity, or by major. Then the Model can identify if the coursework associated with gains in learning among the total group is the same as that for the subgroups. Such appropriate information is valuable to curriculum planners. Courses in the general education sequence not found to be associated with gains in student learning can be revised, enhanced or dropped. Courses outside the general education requirements that contribute to gains in student learning can become candidates for inclusion in the general education curriculum. The extent to which general education courses affect the learning of both high ability and low ability students has relevance in deciding how wide ranging the distributional options should be or whether a core curriculum is appropriate for the students and the educational goals of the institution.

How can the DCP Cluster Analytic Model help with advising students? By linking the

coursework students take with their improvement in learning, the Cluster Analytic Model can be particularly valuable in advising students. First, it takes advising beyond the mere listing of formal degree requirements to the identification of those specific courses in which students of comparable interests, abilities and achievement have enrolled. Given several years of assessment data linked to the transcripts of graduating seniors, the Model can identify an array of courses taken by students who showed the largest gains in general learning in college. The Model is amenable to the development of a microcomputer-based advising system utilizing a relational database of prior students coursetaking patterns and assessment results. Such a computer-based advising system would yield an array of effective coursework tailored to the abilities and interests of individual students and within the parameters of institutional degree requirements. Such a computer-based assessment and transcript system could be built and piloted using a nationally-representative sample of students, such as that in the longitudinal study of the National Center for Postsecondary Teaching, Learning and Assessment.

Conclusions and Summary

We do need a better means of assessing college graduates' abilities to think critically, communicate clearly, and solve problems. For a national program assessing college graduates to succeed, it must be institutionalized at the majority of colleges and universities in the nation. The information generated from such an assessment must play a clear and key role in formulating state and institutional higher education policy, in college matriculation standards and degree-granting decisions. In short, it must affect

the teaching of faculty and the learning behavior of students. For if students don't try to improve their ability to think critically, communicate clearly and solve problems, then the NEGP objective will not be realized. Colleges and universities can make such a national assessment program meaningful to students by making it part of graduation requirements for the baccalaureate and by using the information from the assessment to guide students in making better-informed choices about their educational program.

For colleges and universities to give such meaning to an assessment program, it must be credible, reliable and useful. It must be particularly credible, reliable and useful to faculty, who play a fundamental role in portraying the importance of the assessment program to students. Such confidence can only be developed through the active participation of colleges and universities and their faculty in the development of such an assessment program. Several states and many institutions have substantial investments in currently operating and successful assessment programs. The national effort should build on the credibility and accomplishments of these programs rather than to be duplicative and ancillary to them.

References

Adelman, C. (1988). Opening remarks. "Linking Student Outcomes and Curricular Reform: A New Research Model", a forum presentation at the annual meeting of the American Association for Higher Education (AAHE), Washington, DC, March 9-13, 1988.

Adelman, C. (Ed.). (1986). Assessment in American higher education: Issues and contexts. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

Adelman, C. (1985). The standardized test scores of college graduates, 1964-1982. Washington: Study Group on the Conditions of Excellence in American Higher Education, National Institute of Education.

Borg, W.R. & Gall, M.D. (1983). Educational research: An introduction. Fourth edition. New York: Longman.

Boyer, C.M., & Ahlgren, A. (1987). Assessing undergraduates' patterns of credit distribution: Amount and specialization. Journal of Higher Education, 58, 430-442.

Brubacher J.S., & Rudy, W. (1976). Higher Education in Transition: A History of American Colleges and Universities, 1636-1976. 3rd ed. Harper & Row, New York.

Farland, R. W., & Cepeda, R. (1990). Precollegiate Basic Skills in the California Community Colleges: A Report. Sacramento, CA: California Community Colleges, Office of the Chancellor, 1990. ERIC Document No. ED 317 256.

Gosak, A. (1988). Here Today, Here Tomorrow: A Longitudinal Study of 1,000 Students Entering San Jose City College's ESL Program. San Jose, CA: San Jose City College. ERIC Document No. ED 304 200.

Grant, C.A., & Sleeter, C.E. (1986, Summer). Race, class, and gender in education research: An argument for integrative analysis. Review of Educational Research, 56 (2), 195-211.

Hobbs, R. L. (1989). Academic and Developmental Services End of Year Status Report, 1988-89. Memphis, TN: Shelby State Community College. ERIC Document No. ED 310 835.

King, P.; Wood, P. & Mines, R. (1989). Problem structure, tests of critical thinking and disciplinary differences: A study of critical thinking among college and graduate students. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, 1988.

McCabe, R. H. (1989). Floridians at Risk of Losing Higher Education Opportunity. Miami, FL: Miami-Dade Community College. ERIC Document No. ED 313 087.

National Institute of Education. Study Group on the Conditions of Excellence in American Higher Education. (1984, October). Involvement in learning: Realizing the potential of American higher education. Washington, DC: Government Printing Office.

Pascarella, E. & Terenzini, P. (1991). How College Affects Students: Findings and Insights from Twenty Years of Research. San Francisco: Jossey-Bass.

Ratcliff, J.L. (1990). Development and testing of a cluster-analytic model for identifying coursework patterns associated with general learned abilities of college students: Final report, May 1990 U.S. Department of Education, Office of Educational Research and Improvement, Research Division. Contract No. OERI-R-86-0016. University Park, PA: Center for the Study of Higher Education, Pennsylvania State University.

Reed, R. J. (1987). School and College Competency Testing Programs: Perceptions and Effects on the Black Students in Louisiana and North Carolina. Atlanta, GA: Southern Education Foundation. ERIC Document No. ED 286 960.

Richards, W. (1986). The Effectiveness of New-Student Basic Skills Assessment in Colorado Community Colleges. Denver, CO: Colorado State Board for Community Colleges and Occupational Education, Report prepared by the Assessment and Basic Skills Committee. ERIC Document No. ED 275 351.

Smith, A., & Morris, D. (1989). New issues, problems, programs and research ideas: Findings of a statewide study of Texas community/junior college presidents and chancellors. Paper presented at the Annual Meeting of the American Association of Community and Junior Colleges (69th, Washington, DC, March 29-April 1, 1989). ERIC Document No. 307 007.

Sternberg, R. J. (1989, March). The tyranny of testing. Learning, 17, 60-63.

Sworder, S. (1990). A Review of the Mathematics Assessment of Saddleback College Students through the Matriculation Program. Mission Viejo, CA: Saddleback Community College. ERIC Document No. ED 316 292.

Veysey, L. (1973). Stability and experiment in the American undergraduate curriculum. In C. Kaysen (ed.), Content and Context: Essays on College Education. New York: McGraw-Hill.

Paper: **What Kind of National Assessment Fits American Higher Education?**

Author: James L. Ratcliff, The Pennsylvania State University

Reviewer: Nancy Beck, Educational Testing Service

> This paper focuses on use of one specific model to address the concerns of the Goals Panel in measuring progress toward improved skills, and it assumes familiarity with the *Differential Coursework Patterns Project* and the *Cluster Analytic Model.* Although use of this model receives the greatest emphasis, other instruments/methods are also recommended as part of a proposed assessment system.

The paper is written in support of the particular position and experience of the author's experience with a particular analytic approach, and, while that is not in itself a drawback, in this case that position and single method are being asked to do more than they possibly can.

Although supportive of the concept of some kind of national assessment, the emphasis is on *institutional* use of the information. The author says that the results would be used at the state level, but the focus is entirely on the institution. Use of the information at the national level - in response to the Goal Panel's concerns - really is not addressed.

My understanding of the author's measurement approach is that existing examinations (admissions tests or parts of them) would form the assessment portion of the proposed system. Tests such as the SAT and ACT would be used to get a fix on entering ability; graduate and professional school tests (GRE, LSAT, GMAT, MCAT), or parts of them, would be the exit instruments.

At the core of the measurement, however, would be analysis of course-taking patterns by subgroups within institutions (and states) using the *Differential Coursework Patterns Cluster Analytic Model.* This model would be used to analyze the coursework patterns of subgroups of students and provide useful feedback to an institution so it could identify appropriate and inappropriate learning environments. Institutions would then take action as necessary to change their curricula to encourage course sequences that enhance development of the 5.5 skills of critical thinking, clear communication, and problem solving. Support for this assumption is not provided.

There is no attempt in the paper to discuss the nature of the three skills, or to identify any sub-skills under them (beyond some positive words about the Analytic section of the GRE General Test).

If the task at hand was "...to identify, define, and assess a specific set of skills which are consistent with the stated objective of national goal five...," then, even though the

paper does not directly do that, it is still useful to try to apply the review criteria to the measures Ratcliff proposes.

1. **A valid case was proposed for the measures.** No. While some of the points about impact, acceptance, and student motivation are valid, it is not really clear from the paper that the proposed approach would deal with those and other issues.

2. **Acquisition or possession of the skills can be shown.** Not clear. If the existing measures are accepted as a fair and valid measure of these skills (a large "if"), it may be possible. It is not immediately obvious what the course pattern selection data (which would be expensive to obtain on any kind of broad scale) would contribute here.

3. **Permits identification of growth or value added.** Possibly. If the pre-collegiate admission data and the pre-graduate/professional school admissions data are acceptable as measures of the skills of concern (again the large "if"), then something (of unknown value) could be pulled together for the populations for whom the data are available. It is important to note that this would exclude both community college students and students who do not go on to graduate or professional schools.

4. **Assessments of these skills allows for:**

   - Accurate measurement of each set of skills: Probably not. It is far from clear that the existing instruments could or would generate acceptable assessments of these skills.

   - Determination of barriers to acquisition. Perhaps. This appears to be the heart of Ratcliff's proposal. That is, that analysis of course-taking patterns would permit institutions to identify those course patterns which did not contribute to improved student learning for various groups of students. It was not clear from the paper, however, that this could be done on a wide scale.

- Identification of effective learning environments. Perhaps. This is the other side of the coin - if the system worked to determine barriers, presumably it could also identify effective programs.

5. **Methods are practical, replicable, and complete.**

- Derived from reliable and practical research applications. Not clear. Without knowledge of his report, it is not possible to know whether what is proposed has a solid research basis. The measures used and the data to support them would be interesting, particularly in relation to what the Goals Panel says it wants to measure.

- Adaptable to a national environment or program. Yes, if it were acceptable. That is, since the national tests for entry and exit testing already exist, the problem of would be in the broad implementation of the *Coursework Pattern Model*. Given enough money, it probably could be done on a national basis.

- Requires little or no further research or testing. No. A large amount of additional research would be required to establish that the existing tests could be used to measure what the Panel is interested in (the 5.5 skills). Without reading the *Coursework Patterns Project* report one would have to guess that the proposed Model would also need further research.

- Cost efficient and effective practices. No. Although use of existing tests would be very cost effective, the collection, analysis, and reporting on student coursework patterns and their relation to learning would be a massive, labor-intensive effort. It would be very expensive for what it would yield in terms of the Panel's agenda.

General Comments:

It was difficult to do a satisfactory review of this paper since it did not seem to address the issues raised by the Panel and assumed an understanding of the proposed Model. Another concern that receives litte discussion or documented support is that existing instruments can serve as measures of the 5.5 skills. The fact that the tests cited were not developed as measures of critical thinking, communication, and problem solving is not addressed. Although those skills are, in all probability, measured within them, the paper does not say how they could be extracted to meet the measurement requirements of Goal 5, or how they could deal with the full college entry and college graduate populations.

The *Coursework Pattern* discussion was interesting and appears potentially quite useful on an individual campus; it is not well supported nationally, however, and was not effectively related, in this paper to the Panel's request.

November 1991

Review of
What Type of National Assessment Fits American Higher Education

by
James L. Ratcliff

Review by Joan Herman, UCLA/CRESST

Dr. Ratcliff's paper is very well written, thoughtful, and scholarly; it is based on sound reasoning, shows thorough understanding of the complexities of the higher education context, and supports its positions from a historical perspective as well as from current literature. Unfortunately, however, the paper focuses mostly on the problems of monitoring the process and outcomes of higher education rather than on clarifying the skills/domains and the specific methods by which they might be reasonably assessed.

The paper is very convincing in identifying a range of factors which will impede valid assessment and the productive use of assessment to reach national goals. Among these are the lack of current support among the faculty and institutions of higher ed, whose involvement will be critical; the probable lack of student motivation to do well on the tests if they are not tied to institutional or personal incentives; the difficulty of defining common measures or common standards given the extreme variation in curriculum and educational goals both within and across institutions; potential negative side effects for issues of equity and access; and an overarching concern that standard setting alone will do little or nothing to improve the quality of higher education. The

"national" assessment which Dr. Ratcliff envisions must inform improvement directly. These are all important issues which any national assessment design must address.

In terms of what specifically we should be assessing or how best to assess it, Dr. Ratcliff's paper is relatively mute. He does strongly indicate that currently available means for "evaluating student learning are not satisfactory for a national collegiate assessment," although he does seem to believe that some of the GRE-type measures used in the Differential Coursework Patterns Study have some promise (although he also suggests the need for more research to better understand what they are measuring). Furthermore, he appears to strongly believe in a "value-added" model for assessing goal five, and rightly points out a number of difficulties in validly doing so, among them student attrition and the problems in differentiating the quality of educational programs and their outcomes from differences and/or changes in the student populations served.

As mentioned, Dr. Ratcliff's paper strongly advocates assessment for the purpose of improvement, and it seems clear that he believes that assessment can best fulfill that purpose if it provides information about both the process and outcomes of higher education institutions. This is important advice. Dr. Ratcliff states that the guiding question for a national assessment ought to be "which groups of students benefit most from which collegiate environments." While the answers to such questions are useful to national and other policymakers, he seems most committed to answering such questions in the context of individual institutions,

who can use information on coursework patterns, student outcomes, and student characteristics to fine tune or adjust their educational programs. An implicit and important issue in his paper, then, is how a national assessment system can best be configured to encourage individual institutions to gather and use information about the quality of their processes and outcomes. What are appropriate incentives and structures to support such self-reflection and improvement?

Given the paucity of currently available measures to assess student outcomes, Dr. Ratcliff's paper suggests a two pronged approach which starts with monitoring the quality of coursework opportunities while new measures are researched, developed, and validated. For the former purpose, he advocates analysis of course syllabi and requirements and, later transcript analysis to determine course taking patterns. While this is an interesting interim measure of quality, it also would be a very costly one to implement. If quality of process, however, is thought to be a good short-term proxy for ultimate questions of outcomes, studies at CRESST indicate that self-reports from faculty and students can be an effective and cost efficient strategy for creating that proxy (McDonnell et al, 1990).

Dr. Ratcliff makes the case for multiple measures of outcomes, a critically important requirement for a national assessment system. He also leaves us with an important set of criteria which a meaningful assessment program must meet. Among these are that it is credible, reliable, useful, and fully supported by faculty. But rather than clarifying what we should be assessing, Dr.

Ratcliff leaves us with a research process through which institutions can analyze and use assessment information to improve themselves. To me, this is slightly off-target from the primary question at hand: what types of national and other assessments will help us to monitor progress toward goals, and what is the nature of the skills we should be assessing?.

Review of:

James L. Ratcliff: What Type of National Assessment Fits American Higher Education?

By: Ted Marchese and Barbara Wright, AAHE Assessment Forum

This paper seems to come down beside the point, offering a "solution" -- better advisement -- before dealing adequately with the problem at hand: how best to assess college students' critical thinking, problem solving, and communication skills.

The first nine pages constitute a long argument that the very elements that give rise to outcomes assessment -- poor high schools, diverse students, exploding curricula -- doom the effort to failure. It's an interesting argument, but inconclusively stated. A couple of sentences at the bottom of page 8 and top of page 9 assert the conclusion but without elaboration, discussion, or evidence.

Pages 9-17 consist of many intertwined arguements:

a) tests have to evoke students' best efforts,; for that, tests must have significance to the student and to the institution;

b) gateway tests (like CLAST, TASP) curtail access and do not lead to system-wide improvement or raise student achievement;

c) assessment programs alone, even if perceived as fair and taken seriously, cannot raise achievement;

d) assessment with diagnostic value, used in advisement, is far more likely to heighten achievement;

e) faculty involvement in the process in crucial;

f) most testing programs underrepresent part-time students, working

adults, minorities, commuters;

g) any new test is likely to exhibit bias;

h) without a parallel test at entry, every wrong kind of conclusion is likely to be drawn;

i) tests are likely to restrict access, even as our global competitors seek to expand access;

j) we can't do NAEP-like matrix sampling because curricula in higher education don't have the uniformity of K-12; and

k) differences within institutions are greater than those between institutions (from Pascarella and Terenzini).

This is quite an array of arguments against postsecondary testing; however, the paper does not go on to draw the conclusion that we should therefore consider alternative methods. Comment on each of the eleven isn't possible (most are too briefly stated), but among them, a) and k) caught our attention: a) for its realism and the good reminder, k) for its insight that taking the entire institution as the unit of analysis may be problematic.

Pages 17-20 argue that an assessment must be diagnostic and helpful to effect change -- that is, it has to connect to something fixable. The argument then proposes as appropriate units of analysis 1) programs or sequences of courses; and 2) statewide systems of higher education. In terms of meeting a criterion of "diagnostic/helpful to change," the paper makes clear that there is a choice, but does not clarify why statewide data would be more helpful than, say, institution-level data. Given aggregated statewide data alone, what exactly would a legislature or state board then be in a position to fix? What of institution-level responsibility?

The next section (pages 20-25) seems to argue that a test of general learned abilities should be given yearly to cohorts as they move through

college, in order to identify which patterns of coursework are most beneficial to which groups of students, but notes that more research is needed on all this. Perhaps from the point of view of a researcher, questions of differential impact make sense as a next line of questioning to pursue; from the standpoint of a National Education Goals Panel, however, the appropriate question may be more limited and simpler: Are America's college graduates improving in the three abilities?

The actual proposal (pages 25-29) is in two parts:

a) Develop protocols for syllabi and exams that monitor the extent to which the three abilities appear to be stressed in coursework. The idea is not much explained here, and questions of our readiness to formulate such protocols leap to mind, as do questions about their face validity for faculty, the sheer difficulties of so reviewing the thousands of syllabi and exams generated each year in a university, and so on. In our view, anything that gets teachers to pause a moment over questions of course design, pedagogy, and testing could be a plus. But next we have to ask: how credible would cumulations of such data appear to the public? Not very, we'd guess.

b) Deploy the Differential Coursetaking Patterns model across all institutions. Unfortunately, neither the model itself nor any of its findings are described here, making it impossible for a reader to make an informed judgment of the model's usefulness. We wonder whether higher education is making things too easy for itself if it assumes, as the proposal seems to, that for the most part courses and instruction are not in need of improvement, only students' patterns of course selection. One wonders, too, whether the transcript files of most colleges would support the venture. But giving the proposal the benefit of the doubt on these matters, we ask four further questions:

1) How, exactly, will the three abilities, and students' improvement, be documented? This proposal begs the question. The long, historical introduction, the delineation of the forces that led to creation of the College Board at the beginning of this century, and the assertion that these same forces are affecting higher education today seem to suggest that, with regard to measuring critical thinking skills, what worked then will work now. According to pages 23-24, although more research is needed, the ACT, SAT, and GRE scores provide adequate descriptions of the three abilities. That is the extent of the discussion on this central question.

2) Can a campus-based system in which all the work of assessment is necessarily done by third-party experts (educational researchers), and that makes its case via elaborate computer-generated tables, be persuasive to a broad range of university faculty members? In current practice, the most successful assessment programs have started with faculty questions and had faculty themselves as the inquirer-data gatherers.

3) How persuasive to faculty advisors and to students are regression data, showing "tendencies," at best, from years ago, about a body of courses that is (one would hope) constantly changing and evolving?

4) Can data of this type from dozens of institutions in a state be cumulated in a way that is accurate, credible, and useful to state-level (not to mention national) decision makers? How could they interpret it? What, exactly, would it tell them? What decisions could appropriately follow from it? How would the general public react?